Conventions, Accuracy Metrics, Classification, Regression

Nipun Batra

IIT Gandhinagar

August 3, 2025

Outline

Demo

- Complete PoseNet Demo

Demo

- Complete PoseNet Demo
- Blog post from Google

Demo

- Complete PoseNet Demo
- Blog post from Google
- Rock Paper Scissors

Revision: What is Machine Learning

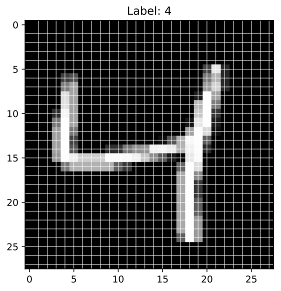
"Field of study that gives computers the ability to learn without being explicitly programmed" - Arthur Samuel [1959]

Revision: What is Machine Learning

"Field of study that gives computers the ability to learn without being explicitly programmed" - Arthur Samuel [1959]

Let us work on the digit recognition problem.

Notebook: rule-based-vs-ml.html



Revision: What is Machine Learning

- How would you program to recognise digits? Start with 4.

Revision: What is Machine Learning

- How would you program to recognise digits? Start with 4.

Revision: What is Machine Learning

- How would you program to recognise digits? Start with 4.

Revision: What is Machine Learning

- How would you program to recognise digits? Start with 4.
- Maybe 4 can be thought of as: | + — + | + another vertically down |

Revision: What is Machine Learning

- How would you program to recognise digits? Start with 4.
- Maybe 4 can be thought of as: | + — + | + another vertically down |

Revision: What is Machine Learning

- How would you program to recognise digits? Start with 4.
- Maybe 4 can be thought of as: | + — + | + another vertically down |
- The heights of each of the | need to be similar within tolerance

Revision: What is Machine Learning

- How would you program to recognise digits? Start with 4.
- Maybe 4 can be thought of as: | + — + | + another vertically down |
- The heights of each of the | need to be similar within tolerance

Revision: What is Machine Learning

- How would you program to recognise digits? Start with 4.
- Maybe 4 can be thought of as: | + — + | + another vertically down |
- The heights of each of the | need to be similar within tolerance
- Each of the | can be slightly slanted. Similarly the horizontal line can be slanted.

Revision: What is Machine Learning

- How would you program to recognise digits? Start with 4.
- Maybe 4 can be thought of as: | + — + | + another vertically down |
- The heights of each of the | need to be similar within tolerance
- Each of the | can be slightly slanted. Similarly the horizontal line can be slanted.

Revision: What is Machine Learning

- How would you program to recognise digits? Start with 4.
- Maybe 4 can be thought of as: | + — + | + another vertically down |
- The heights of each of the | need to be similar within tolerance
- Each of the | can be slightly slanted. Similarly the horizontal line can be slanted.
- There can be some cases of 4 where the first | is at 45 degrees

Revision: What is Machine Learning

- How would you program to recognise digits? Start with 4.
- Maybe 4 can be thought of as: | + — + | + another vertically down |
- The heights of each of the | need to be similar within tolerance
- Each of the | can be slightly slanted. Similarly the horizontal line can be slanted.
- There can be some cases of 4 where the first | is at 45 degrees

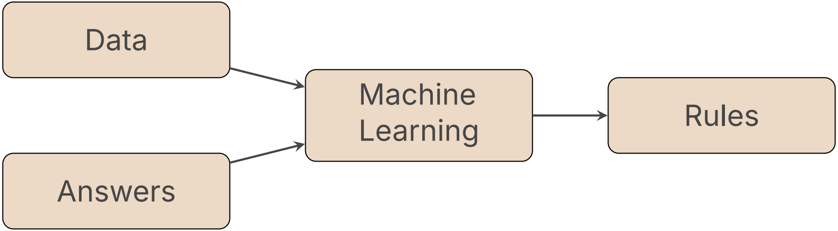
Revision: What is Machine Learning

- How would you program to recognise digits? Start with 4.
- Maybe 4 can be thought of as: | + — + | + another vertically down |
- The heights of each of the | need to be similar within tolerance
- Each of the | can be slightly slanted. Similarly the horizontal line can be slanted.
- There can be some cases of 4 where the first | is at 45 degrees
- There can be some cases of 4 where the width of each stroke is different

Traditional Programming vs Machine Learning

Traditional Programming

Machine Learning



Revision: What is Machine Learning

"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E." - Tom Mitchell

Pop Quiz #1

Quick Quiz 1

In machine learning, which of the following is typically NOT a useful feature?

- a) Color of a tomato for quality prediction

Pop Quiz #1

Quick Quiz 1

In machine learning, which of the following is typically NOT a useful feature?

- a) Color of a tomato for quality prediction
- b) Size of a house for price prediction

Pop Quiz #1

Quick Quiz 1

In machine learning, which of the following is typically NOT a useful feature?

- a) Color of a tomato for quality prediction
- b) Size of a house for price prediction
- c) Sample ID number

Pop Quiz #1

Quick Quiz 1

In machine learning, which of the following is typically NOT a useful feature?

- a) Color of a tomato for quality prediction
- b) Size of a house for price prediction
- c) Sample ID number
- d) Age for medical diagnosis

Pop Quiz #1 - Answer

Answer: c) Sample ID numbers are arbitrary identifiers, not meaningful features!

First ML Task: Grocery Store Tomato Quality Prediction

Problem statement: You want to predict the quality of a tomato given its visual features.

Dataset

Imagine you have some past data on quality of tomatoes.
What visual features do you think will be useful?

Dataset

Imagine you have some past data on quality of tomatoes.
What visual features do you think will be useful?

Dataset

Imagine you have some past data on quality of tomatoes.
What visual features do you think will be useful?

- Size

Dataset

Imagine you have some past data on quality of tomatoes.
What visual features do you think will be useful?

- Size

Dataset

Imagine you have some past data on quality of tomatoes.
What visual features do you think will be useful?

- Size
- Colour

Dataset

Imagine you have some past data on quality of tomatoes.
What visual features do you think will be useful?

- Size
- Colour

Dataset

Imagine you have some past data on quality of tomatoes.
What visual features do you think will be useful?

- Size
- Colour
- Texture

Sample Dataset

Here is our example dataset with tomato features:

Sample	Colour	Size	Texture	Condition
1	Orange	Small	Smooth	Good
2	Red	Small	Rough	Good
3	Orange	Medium	Smooth	Bad
4	Yellow	Large	Smooth	Bad

Useful Features

Is the sample number a useful feature for predicting quality of a tomato?

Useful Features

Is the sample number a useful feature for predicting quality of a tomato?
Answer: Usually no! Sample numbers are typically arbitrary identifiers and not meaningful features. Let us remove it.

Useful Features

Is the sample number a useful feature for predicting quality of a tomato?
Answer: Usually no! Sample numbers are typically arbitrary identifiers and not meaningful features. Let us remove it.
Let us modify our data table for now.

Colour	**Size**	**Texture**	**Condition**
Orange	Small	Smooth	Good
Red	Small	Rough	Good
Orange	Medium	Smooth	Bad
Yellow	Large	Smooth	Bad

Training Set

Colour	Size	Texture	Condition
Orange	Small	Smooth	Good
Red	Small	Rough	Good
Orange	Medium	Smooth	Bad
Yellow	Large	Smooth	Bad

Training Set

Colour	**Size**	**Texture**	**Condition**
Orange	Small	Smooth	Good
Red	Small	Rough	Good
Orange	Medium	Smooth	Bad
Yellow	Large	Smooth	Bad

The training set consists of two parts:

Training Set

Colour	**Size**	**Texture**	**Condition**
Orange	Small	Smooth	Good
Red	Small	Rough	Good
Orange	Medium	Smooth	Bad
Yellow	Large	Smooth	Bad

The training set consists of two parts:

Training Set

Colour	Size	Texture	Condition
Orange	Small	Smooth	Good
Red	Small	Rough	Good
Orange	Medium	Smooth	Bad
Yellow	Large	Smooth	Bad

The training set consists of two parts:

1. Features (Input Variables)

Training Set

Colour	Size	Texture	Condition
Orange	Small	Smooth	Good
Red	Small	Rough	Good
Orange	Medium	Smooth	Bad
Yellow	Large	Smooth	Bad

The training set consists of two parts:

1. Features (Input Variables)

Training Set

Colour	Size	Texture	Condition
Orange	Small	Smooth	Good
Red	Small	Rough	Good
Orange	Medium	Smooth	Bad
Yellow	Large	Smooth	Bad

The training set consists of two parts:

1. Features (Input Variables)
2. Output or Response Variable

Training Set

Colour	Size	Texture	Condition
Orange	Small	Smooth	Good
Red	Small	Rough	Good
Orange	Medium	Smooth	Bad
Yellow	Large	Smooth	Bad

Training Set

Colour	Size	Texture	Condition
Orange	Small	Smooth	Good
Red	Small	Rough	Good
Orange	Medium	Smooth	Bad
Yellow	Large	Smooth	Bad

We call this matrix as $\mathcal{D}$, containing:

1. Feature matrix ($\mathbf{X} \in \mathbb{R}^{n \times d}$) containing data of n samples each of which is d dimensional.

Training Set

Colour	Size	Texture	Condition
Orange	Small	Smooth	Good
Red	Small	Rough	Good
Orange	Medium	Smooth	Bad
Yellow	Large	Smooth	Bad

We call this matrix as $\mathcal{D}$, containing:

1. Feature matrix ($\mathbf{X} \in \mathbb{R}^{n \times d}$) containing data of n samples each of which is d dimensional.
2. Output vector ($\mathbf{y} \in \mathbb{R}^{n}$) containing output variable for n samples.

Data Matrix Details

- Feature matrix: $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \\ \vdots \\ \mathbf{x}_n^\top \end{bmatrix}$ where $\mathbf{x}_i \in \mathbb{R}^d$

Data Matrix Details

- Feature matrix: $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \\ \vdots \\ \mathbf{x}_n^\top \end{bmatrix}$ where $\mathbf{x}_i \in \mathbb{R}^d$

Data Matrix Details

- Feature matrix: $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \\ \vdots \\ \mathbf{x}_n^\top \end{bmatrix}$ where $\mathbf{x}_i \in \mathbb{R}^d$

Data Matrix Details

- Feature matrix: $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \\ \vdots \\ \mathbf{x}_n^\top \end{bmatrix}$ where $\mathbf{x}_i \in \mathbb{R}^d$
- Example (after encoding): $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$ (Orange=1, Small=0, Smooth=1)

Data Matrix Details

- Feature matrix: $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \\ \vdots \\ \mathbf{x}_n^\top \end{bmatrix}$ where $\mathbf{x}_i \in \mathbb{R}^d$
- Example (after encoding): $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$ (Orange=1, Small=0, Smooth=1)

Data Matrix Details

- Feature matrix: $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \\ \vdots \\ \mathbf{x}_n^\top \end{bmatrix}$ where $\mathbf{x}_i \in \mathbb{R}^d$
- Example (after encoding): $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$ (Orange=1, Small=0, Smooth=1)
- Complete dataset: $\mathcal{D} = \{(\mathbf{x}_i^\top, y_i)\}_{i=1}^n$

Prediction Task

Estimate condition for unseen tomatoes (#5, 6) based on data set.

Colour	**Size**	**Texture**	**Condition**
Orange	Small	Smooth	Good
Red	Small	Rough	Good
Orange	Medium	Smooth	Bad
Yellow	Large	Smooth	Bad
Red	Large	Rough	?
Orange	Large	Rough	?

Testing Set

Testing set is similar to training set, but, does not contain labels for output variable.

Colour	Size	Texture	Condition
Orange	Small	Smooth	Good
Red	Small	Rough	Good
Orange	Medium	Smooth	Bad
Yellow	Large	Smooth	Bad
Red	Large	Rough	?
Orange	Large	Rough	?

Prediction Task

We hope to:

Prediction Task

We hope to:

Prediction Task

We hope to:

1. Learn f: Condition $= f(\text{colour, size, texture})$

Prediction Task

We hope to:

1. Learn f: Condition $= f(\text{colour, size, texture})$

Prediction Task

We hope to:

1. Learn f: Condition $= f($colour, size, texture$)$
2. From Training Dataset

Prediction Task

We hope to:

1. Learn f: Condition $= f(\text{colour, size, texture})$
2. From Training Dataset

Prediction Task

We hope to:

1. Learn f: Condition $= f($colour, size, texture$)$
2. From Training Dataset
3. To Predict the condition for the Testing set

Colour	**Size**	**Texture**	**Condition**
Orange	Small	Smooth	Good
Red	Small	Rough	Good
Orange	Medium	Smooth	Bad
Yellow	Large	Smooth	Bad
Red	Large	Rough	?
Orange	Large	Rough	?

Generalisation

- Q: Is predicting on test set enough to say our model generalises?

Generalisation

- Q: Is predicting on test set enough to say our model generalises?

Generalisation

- Q: Is predicting on test set enough to say our model generalises?

Generalisation

- Q: Is predicting on test set enough to say our model generalises?
- A: Ideally, no!

Generalisation

- Q: Is predicting on test set enough to say our model generalises?
- A: Ideally, no!

Generalisation

- Q: Is predicting on test set enough to say our model generalises?
- A: Ideally, no!
- Ideally - we want to predict “well” on all possible inputs. But, can we test that?

Generalisation

- Q: Is predicting on test set enough to say our model generalises?
- A: Ideally, no!
- Ideally - we want to predict "well" on all possible inputs. But, can we test that?

Generalisation

- Q: Is predicting on test set enough to say our model generalises?
- A: Ideally, no!
- Ideally - we want to predict "well" on all possible inputs. But, can we test that?
- No! Since, the test set is only a sample from all possible inputs.

Generalisation

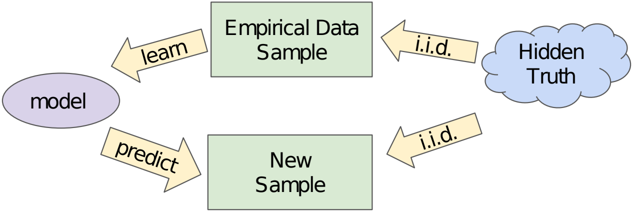


Image courtesy Google ML crash course

Generalisation

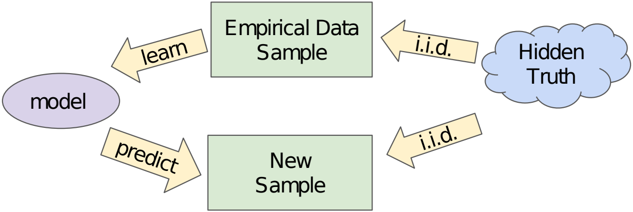


Image courtesy Google ML crash course

Both the training set and the test set are samples drawn from the hidden true distribution (also sometimes called population)

Generalisation

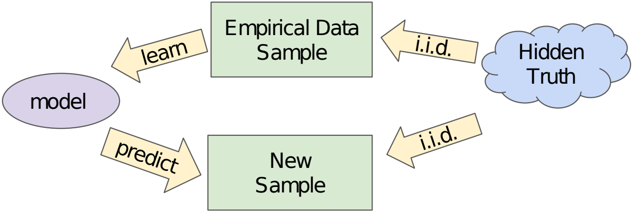


Image courtesy Google ML crash course

Both the training set and the test set are samples drawn from the hidden true distribution (also sometimes called population)
More discussion later once we study bias and variance

Second ML Task: Predict energy consumption of campus

Question: What factors does the campus energy consumption depend on?
Answer:

Second ML Task: Predict energy consumption of campus

Question: What factors does the campus energy consumption depend on?
Answer:

Second ML Task: Predict energy consumption of campus

Question: What factors does the campus energy consumption depend on?
Answer:

- # People (More people $\implies$ More Energy)

Second ML Task: Predict energy consumption of campus

Question: What factors does the campus energy consumption depend on?
Answer:

- # People (More people $\implies$ More Energy)

Second ML Task: Predict energy consumption of campus

Question: What factors does the campus energy consumption depend on?
Answer:

- # People (More people $\implies$ More Energy)
- Temperature (Higher Temp. $\implies$ Higher Energy)

Second ML Task: Predict energy consumption of campus

Question: What factors does the campus energy consumption depend on?
Answer:

- # People (More people $\implies$ More Energy)
- Temperature (Higher Temp. $\implies$ Higher Energy)

# People	Temp (C)	Energy (kWh)
4000	30	30
4200	30	32
4200	35	40
3000	20	?
1000	45	?

Pop Quiz #2

Quick Quiz 2

Which of these is a regression problem?

- a) Predicting if an email is spam or not

Pop Quiz #2

Quick Quiz 2

Which of these is a regression problem?

- a) Predicting if an email is spam or not
- b) Classifying images as cat, dog, or bird

Pop Quiz #2

Quick Quiz 2

Which of these is a regression problem?

- a) Predicting if an email is spam or not
- b) Classifying images as cat, dog, or bird
- c) Predicting house prices

Pop Quiz #2

Quick Quiz 2

Which of these is a regression problem?

- a) Predicting if an email is spam or not
- b) Classifying images as cat, dog, or bird
- c) Predicting house prices
- d) Determining if a tumor is malignant or benign

Pop Quiz #2 - Answer

Answer: c) House prices are continuous values - that's regression!

Classification vs Regression

- Classification

Classification vs Regression

- Classification

Classification vs Regression

- Classification

Classification vs Regression

- Classification
    - Output variable is discrete

Classification vs Regression

- Classification
    - Output variable is discrete

Classification vs Regression

- Classification
  - Output variable is discrete
  - i.e. $y_i \in \{1, 2, \ldots, k\}$ where k is number of classes

Classification vs Regression

- Classification
  - Output variable is discrete
  - i.e. $y_i \in \{1, 2, \ldots, k\}$ where k is number of classes

Classification vs Regression

- Classification
  - Output variable is discrete
  - i.e. $y_i \in \{1, 2, \ldots, k\}$ where k is number of classes
  - Examples - Predicting:

Classification vs Regression

- Classification
  - Output variable is discrete
  - i.e. $y_i \in \{1, 2, \ldots, k\}$ where k is number of classes
  - Examples - Predicting:

Classification vs Regression

- Classification
  - Output variable is discrete
  - i.e. $y_i \in \{1, 2, \ldots, k\}$ where k is number of classes
  - Examples - Predicting:
    - Will I get a loan? (Yes, No)

Classification vs Regression

- Classification
  - Output variable is discrete
  - i.e. $y_i \in \{1, 2, \ldots, k\}$ where k is number of classes
  - Examples - Predicting:
    - Will I get a loan? (Yes, No)

Classification vs Regression

- Classification
  - Output variable is discrete
  - i.e. $y_i \in \{1, 2, \ldots, k\}$ where k is number of classes
  - Examples - Predicting:
    - Will I get a loan? (Yes, No)
    - What is the quality of fruit? (Good, Bad)

Classification vs Regression

- Classification
  - Output variable is discrete
  - i.e. $y_i \in \{1, 2, \ldots, k\}$ where k is number of classes
  - Examples - Predicting:
    - Will I get a loan? (Yes, No)
    - What is the quality of fruit? (Good, Bad)

Classification vs Regression

- Classification
  - Output variable is discrete
  - i.e. $y_i \in \{1, 2, \ldots, k\}$ where k is number of classes
  - Examples - Predicting:
    - Will I get a loan? (Yes, No)
    - What is the quality of fruit? (Good, Bad)
- Regression

Classification vs Regression

- Classification
  - Output variable is discrete
  - i.e. $y_i \in \{1, 2, \ldots, k\}$ where k is number of classes
  - Examples - Predicting:
    - Will I get a loan? (Yes, No)
    - What is the quality of fruit? (Good, Bad)
- Regression

Classification vs Regression

- Classification
  - Output variable is discrete
  - i.e. $y_i \in \{1, 2, \ldots, k\}$ where k is number of classes
  - Examples - Predicting:
    - Will I get a loan? (Yes, No)
    - What is the quality of fruit? (Good, Bad)
- Regression
  - Output variable is continuous

Classification vs Regression

- Classification
  - Output variable is discrete
  - i.e. $y_i \in \{1, 2, \ldots, k\}$ where k is number of classes
  - Examples - Predicting:
    - Will I get a loan? (Yes, No)
    - What is the quality of fruit? (Good, Bad)
- Regression
  - Output variable is continuous

Classification vs Regression

- Classification
  - Output variable is discrete
  - i.e. $y_i \in \{1, 2, \ldots, k\}$ where k is number of classes
  - Examples - Predicting:
    - Will I get a loan? (Yes, No)
    - What is the quality of fruit? (Good, Bad)
- Regression
  - Output variable is continuous
  - i.e. $y_i \in \mathbb{R}$

Classification vs Regression

- Classification
  - Output variable is discrete
  - i.e. $y_i \in \{1, 2, \ldots, k\}$ where k is number of classes
  - Examples - Predicting:
    - Will I get a loan? (Yes, No)
    - What is the quality of fruit? (Good, Bad)
- Regression
  - Output variable is continuous
  - i.e. $y_i \in \mathbb{R}$

Classification vs Regression

- Classification
  - Output variable is discrete
  - i.e. $y_i \in \{1, 2, \ldots, k\}$ where k is number of classes
  - Examples - Predicting:
    - Will I get a loan? (Yes, No)
    - What is the quality of fruit? (Good, Bad)
- Regression
  - Output variable is continuous
  - i.e. $y_i \in \mathbb{R}$
  - Examples - Predicting:

Classification vs Regression

- Classification
  - Output variable is discrete
  - i.e. $y_i \in \{1, 2, \ldots, k\}$ where k is number of classes
  - Examples - Predicting:
    - Will I get a loan? (Yes, No)
    - What is the quality of fruit? (Good, Bad)
- Regression
  - Output variable is continuous
  - i.e. $y_i \in \mathbb{R}$
  - Examples - Predicting:

Classification vs Regression

- Classification
  - Output variable is discrete
  - i.e. $y_i \in \{1, 2, \ldots, k\}$ where k is number of classes
  - Examples - Predicting:
    - Will I get a loan? (Yes, No)
    - What is the quality of fruit? (Good, Bad)
- Regression
  - Output variable is continuous
  - i.e. $y_i \in \mathbb{R}$
  - Examples - Predicting:
    - How much energy will campus consume?

Classification vs Regression

- Classification
  - Output variable is discrete
  - i.e. $y_i \in \{1, 2, \ldots, k\}$ where k is number of classes
  - Examples - Predicting:
    - Will I get a loan? (Yes, No)
    - What is the quality of fruit? (Good, Bad)
- Regression
  - Output variable is continuous
  - i.e. $y_i \in \mathbb{R}$
  - Examples - Predicting:
    - How much energy will campus consume?

Classification vs Regression

- Classification
  - Output variable is discrete
  - i.e. $y_i \in \{1, 2, \ldots, k\}$ where k is number of classes
  - Examples - Predicting:
    - Will I get a loan? (Yes, No)
    - What is the quality of fruit? (Good, Bad)
- Regression
  - Output variable is continuous
  - i.e. $y_i \in \mathbb{R}$
  - Examples - Predicting:
    - How much energy will campus consume?
    - How much rainfall will fall?

Metrics for Classification

$$\text{Prediction } (\hat{y}) \quad \begin{pmatrix} \text{Good} \\ \text{Good} \\ \text{Good} \\ \text{Good} \\ \text{Bad} \end{pmatrix} \qquad \text{Ground Truth } (\mathbf{y}) \quad \begin{pmatrix} \text{Good} \\ \text{Good} \\ \text{Bad} \\ \text{Bad} \\ \text{Bad} \end{pmatrix}$$

Ground Truth:	From the actual training set
Prediction:	Made by the model

Accuracy

$$
\begin{array}{cc}
 & \text{Prediction } (\hat{y}) \\
\begin{array}{c} \checkmark \\ \checkmark \\ \\ \\ \checkmark \end{array} & \begin{pmatrix} \text{Good} \\ \text{Good} \\ \text{Good} \\ \text{Good} \\ \text{Bad} \end{pmatrix}
\end{array}
\qquad\qquad
\begin{array}{c}
\text{Ground Truth } (\mathbf{y}) \\
\begin{pmatrix} \text{Good} \\ \text{Good} \\ \text{Bad} \\ \text{Bad} \\ \text{Bad} \end{pmatrix}
\end{array}
$$

Accuracy

$$\begin{array}{c c c} & \text{Prediction } (\hat{y}) & & \text{Ground Truth } (\mathbf{y}) \\ \begin{array}{c}\checkmark\\ \checkmark\\ \\ \\ \checkmark\end{array} & \begin{pmatrix} \text{Good} \\ \text{Good} \\ \text{Good} \\ \text{Good} \\ \text{Bad} \end{pmatrix} & & \begin{pmatrix} \text{Good} \\ \text{Good} \\ \text{Bad} \\ \text{Bad} \\ \text{Bad} \end{pmatrix} \end{array}$$

$$\begin{aligned} \text{Accuracy} &= \frac{|\{i : y_i = \hat{y}_i\}|}{n} \\ &= \frac{3}{5} = 0.6 \end{aligned}$$

Mathematical Notation: Set Cardinality and Indicator Functions

Mathematical Notation: Set Cardinality and Indicator Functions

Mathematical Notation: Set Cardinality and Indicator Functions

- **Set cardinality notation:** $|\{i : y_i = \hat{y}_i\}|$
  - Reads as: "Number of indices i such that $y_i = \hat{y}_i$"

Mathematical Notation: Set Cardinality and Indicator Functions

- **Set cardinality notation:** $|\{i : y_i = \hat{y}_i\}|$
  - Reads as: "Number of indices i such that $y_i = \hat{y}_i$"
  - Counts how many samples satisfy the condition

Mathematical Notation: Set Cardinality and Indicator Functions

- **Set cardinality notation:** $|\{i : y_i = \hat{y}_i\}|$
  - Reads as: "Number of indices i such that $y_i = \hat{y}_i$"
  - Counts how many samples satisfy the condition

Mathematical Notation: Set Cardinality and Indicator Functions

- **Set cardinality notation:** $|\{i : y_i = \hat{y}_i\}|$
  - Reads as: "Number of indices i such that $y_i = \hat{y}_i$"
  - Counts how many samples satisfy the condition

Mathematical Notation: Set Cardinality and Indicator Functions

- **Set cardinality notation:** $|\{i : y_i = \hat{y}_i\}|$
  - Reads as: "Number of indices i such that $y_i = \hat{y}_i$"
  - Counts how many samples satisfy the condition
- **Alternative: Indicator function notation**

$$\text{Accuracy} = \frac{\sum_{i=1}^{n} \mathbf{1}[y_i = \hat{y}_i]}{n}$$

$$\text{where } \mathbf{1}[\text{condition}] = \begin{cases} 1 & \text{if condition is true} \\ 0 & \text{if condition is false} \end{cases}$$

Mathematical Notation: Set Cardinality and Indicator Functions

- **Set cardinality notation:** $|\{i : y_i = \hat{y}_i\}|$
  - Reads as: "Number of indices i such that $y_i = \hat{y}_i$"
  - Counts how many samples satisfy the condition
- **Alternative: Indicator function notation**

$$\text{Accuracy} = \frac{\sum_{i=1}^{n} \mathbf{1}[y_i = \hat{y}_i]}{n}$$

$$\text{where } \mathbf{1}[\text{condition}] = \begin{cases} 1 & \text{if condition is true} \\ 0 & \text{if condition is false} \end{cases}$$

Mathematical Notation: Set Cardinality and Indicator Functions

- **Set cardinality notation:** $|\{i : y_i = \hat{y}_i\}|$
  - Reads as: "Number of indices i such that $y_i = \hat{y}_i$"
  - Counts how many samples satisfy the condition
- **Alternative: Indicator function notation**

$$\text{Accuracy} = \frac{\sum_{i=1}^{n} \mathbf{1}[y_i = \hat{y}_i]}{n}$$

where $\mathbf{1}[\text{condition}] = \begin{cases} 1 & \text{if condition is true} \\ 0 & \text{if condition is false} \end{cases}$

- Both notations are mathematically equivalent and commonly used in ML literature

Types of Data: Imbalanced Classes

$$
\begin{array}{r}
\text{1 sample} \left\{ \vphantom{\text{Bad}} \right. \\
\text{100 samples} \left\{ \vphantom{\begin{array}{c}\text{Good}\\\text{Good}\\\cdots\\\text{Good}\end{array}} \right.
\end{array}
\begin{pmatrix} \text{Bad} \\ \text{Good} \\ \text{Good} \\ \cdots \\ \text{Good} \end{pmatrix}
$$

Imbalanced Classes

Types of Data: Imbalanced Classes

$$\begin{array}{r}\text{1 sample}\left\{\begin{array}{l}\\\end{array}\right.\\ \text{100 samples}\left\{\begin{array}{l}\\ \\ \\ \\ \end{array}\right.\end{array}\begin{pmatrix}\text{Bad}\\ \text{Good}\\ \text{Good}\\ \dots\\ \text{Good}\end{pmatrix}\qquad\qquad \text{Imbalanced Classes}$$

Cases for this:

- Cancer Screening

Types of Data: Imbalanced Classes

$$\begin{array}{r} \text{1 sample} \{ \\ \text{100 samples} \left\{ \begin{array}{c} \\ \\ \\ \\ \end{array} \right. \end{array} \begin{pmatrix} \text{Bad} \\ \text{Good} \\ \text{Good} \\ \dots \\ \text{Good} \end{pmatrix} \qquad \text{Imbalanced Classes}$$

Cases for this:

- Cancer Screening
- Planet Detection

Accuracy Metrics: Precision

$$
\begin{array}{cc}
 & \text{Prediction } (\hat{y}) \\
\rightarrow \checkmark & \\
\rightarrow \checkmark & \\
\rightarrow & \\
\rightarrow & \\
 &
\end{array}
\begin{pmatrix} \text{Good} \\ \text{Good} \\ \text{Good} \\ \text{Good} \\ \text{Bad} \end{pmatrix}
\qquad
\begin{array}{c} \text{Ground Truth } (\mathbf{y}) \end{array}
\begin{pmatrix} \text{Good} \\ \text{Good} \\ \text{Bad} \\ \text{Bad} \\ \text{Good} \end{pmatrix}
$$

$$\text{Precision} = \frac{|\{i : y_i = \hat{y}_i = \text{Good}\}|}{|\{i : \hat{y}_i = \text{Good}\}|} = \frac{2}{4} = 0.5$$

"the fraction of relevant instances among the retrieved instances", i.e. "out of the number of times we predict Good, how many times is the condition actually Good"

Accuracy Metrics: Precision

$$
\begin{array}{cc}
 & \text{Prediction } (\hat{y}) \\
\begin{array}{c} \rightarrow \checkmark \\ \rightarrow \checkmark \\ \rightarrow \\ \rightarrow \\ \end{array} & \begin{pmatrix} \text{Good} \\ \text{Good} \\ \text{Good} \\ \text{Good} \\ \text{Bad} \end{pmatrix}
\end{array}
\qquad
\begin{array}{c}
\text{Ground Truth } (\mathbf{y}) \\
\begin{pmatrix} \text{Good} \\ \text{Good} \\ \text{Bad} \\ \text{Bad} \\ \text{Good} \end{pmatrix}
\end{array}
$$

$$\text{Precision} = \frac{|\{i : y_i = \hat{y}_i = \text{Good}\}|}{|\{i : \hat{y}_i = \text{Good}\}|} = \frac{2}{4} = 0.5$$

"the fraction of relevant instances among the retrieved instances", i.e. "out of the number of times we predict Good, how many times is the condition actually Good"

Accuracy Metrics: Recall

$$
\begin{array}{ll}
 & \text{Prediction } (\hat{y}) \\
\begin{array}{l} \rightarrow \checkmark \\ \rightarrow \checkmark \\ \\ \\ \rightarrow \end{array} & \begin{pmatrix} \text{Good} \\ \text{Good} \\ \text{Good} \\ \text{Good} \\ \text{Bad} \end{pmatrix}
\end{array}
\qquad
\begin{array}{c}
\text{Ground Truth } (\mathbf{y}) \\
\begin{pmatrix} \text{Good} \\ \text{Good} \\ \text{Bad} \\ \text{Bad} \\ \text{Good} \end{pmatrix}
\end{array}
$$

$$
\text{Recall} = \frac{|\{i : y_i = \hat{y}_i = \text{Good}\}|}{|\{i : y_i = \text{Good}\}|} = \frac{2}{3} = 0.67
$$

"the fraction of the total amount of relevant instances that were actually retrieved"

Pop Quiz #3

Quick Quiz 3

In a dataset of 1000 samples where only 10 are positive cases, what's the accuracy of a classifier that always predicts "negative"?

- a) 1%

Pop Quiz #3

Quick Quiz 3

In a dataset of 1000 samples where only 10 are positive cases, what's the accuracy of a classifier that always predicts "negative"?

- a) 1%
- b) 50%

Pop Quiz #3

Quick Quiz 3

In a dataset of 1000 samples where only 10 are positive cases, what's the accuracy of a classifier that always predicts "negative"?

- a) 1%
- b) 50%
- c) 90%

Pop Quiz #3

Quick Quiz 3

In a dataset of 1000 samples where only 10 are positive cases, what's the accuracy of a classifier that always predicts "negative"?

- a) 1%
- b) 50%
- c) 90%
- d) 99%

Pop Quiz #3 - Answer

Answer: d) 99% - but this classifier is useless! This shows why accuracy can be misleading.

Types of Data: Imbalanced Classes

Given predictions of whether a tissue is cancerous or not ($n = 100$).

$$\text{Prediction } (\hat{y}) \qquad\qquad \text{Ground Truth } (\mathbf{y})$$

$$\begin{array}{r}\rightarrow\\ \\ \\ \\ \\ \end{array}\begin{pmatrix} \text{Yes} \\ \text{No} \\ \text{No} \\ \cdots \\ \text{No} \end{pmatrix} \qquad \begin{array}{r} \\ \\ \\ \\ \rightarrow \end{array}\begin{pmatrix} \text{No} \\ \text{No} \\ \cdots \\ \text{No} \\ \text{Yes} \end{pmatrix}$$

Types of Data: Imbalanced Classes

Given predictions of whether a tissue is cancerous or not ($n = 100$).

$$
\begin{array}{cc}
\text{Prediction } (\hat{y}) & \text{Ground Truth } (\mathbf{y}) \\
\begin{array}{r}\rightarrow \\ \\ \\ \\ \\ \end{array}\begin{pmatrix} \text{Yes} \\ \text{No} \\ \text{No} \\ \dots \\ \text{No} \end{pmatrix} & \begin{array}{r} \\ \\ \\ \\ \rightarrow \end{array}\begin{pmatrix} \text{No} \\ \text{No} \\ \dots \\ \text{No} \\ \text{Yes} \end{pmatrix}
\end{array}
$$

$$\text{Accuracy} = \frac{98}{100} = 0.98$$

$$\text{Recall} = \frac{0}{1} = 0$$

$$\text{Precision} = \frac{0}{1} = 0$$

Accuracy Metrics: Confusion Matrix

		Ground Truth	
		Yes	No
Predicted	Yes	0	1
	No	1	98

Accuracy Metrics: Confusion Matrix

		Ground Truth	
		Yes	No
Predicted	Yes	0	1
	No	1	98

		Ground Truth	
		Yes	No
Predicted	Yes	True Positive	False Positive
	No	False Negative	True Negative

Accuracy Metric: Confusion Matrix

		Ground Truth	
		Yes	No
Predicted	Yes	True Positive	False Positive
	No	False Negative	True Negative

$$\text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}}$$

Accuracy Metric: Confusion Matrix

		Ground Truth	
		Yes	No
Predicted	Yes	True Positive	False Positive
	No	False Negative	True Negative

Precision = $\frac{TP}{TP+FP}$

Accuracy Metric: Confusion Matrix

		Ground Truth	
		Yes	No
Predicted	Yes	True Positive	False Positive
	No	False Negative	True Negative

Recall = $\frac{\text{TP}}{\text{TP}+\text{FN}}$

Accuracy Metric: Confusion Matrix

		Ground Truth	
		Yes	No
Predicted	Yes	True Positive	False Positive
	No	False Negative	True Negative

$$\text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FN}}$$

Enhanced Confusion Matrix: Complete Picture

Confusion Matrix		Actual (Ground Truth)	
		Positive	Negative
Predicted	Positive	TP	FP
	Negative	FN	TN

Definition: Confusion Matrix Elements

- **TP (True Positive):** Correctly predicted positive cases

Enhanced Confusion Matrix: Complete Picture

Confusion Matrix		Actual (Ground Truth)	
		Positive	Negative
Predicted	Positive	TP	FP
	Negative	FN	TN

Definition: Confusion Matrix Elements

- **TP (True Positive):** Correctly predicted positive cases
- **TN (True Negative):** Correctly predicted negative cases

Enhanced Confusion Matrix: Complete Picture

Confusion Matrix		Actual (Ground Truth)	
		Positive	Negative
Predicted	Positive	TP	FP
	Negative	FN	TN

Definition: Confusion Matrix Elements

- **TP (True Positive):** Correctly predicted positive cases
- **TN (True Negative):** Correctly predicted negative cases
- **FP (False Positive):** Wrongly predicted as positive (Type I error)

Enhanced Confusion Matrix: Complete Picture

Confusion Matrix		Actual (Ground Truth)	
		Positive	Negative
Predicted	Positive	TP	FP
	Negative	FN	TN

Definition: Confusion Matrix Elements

- **TP (True Positive):** Correctly predicted positive cases
- **TN (True Negative):** Correctly predicted negative cases
- **FP (False Positive):** Wrongly predicted as positive (Type I error)

Enhanced Confusion Matrix: Precision Focus

Confusion Matrix		Actual		Totals
		Positive	Negative	
Predicted	Positive	TP	FP	TP + FP
	Negative	FN	TN	FN + TN
Totals		TP + FN	FP + TN	All

Example: Precision = Focus on Predicted Positives

$$\text{Precision} = \frac{\textbf{TP}}{\textbf{TP + FP}} = \frac{\text{Correctly predicted positives}}{\text{All predicted positives}}$$

Enhanced Confusion Matrix: Recall Focus

Confusion Matrix		Actual		Totals
		Positive	Negative	
Predicted	Positive	TP	FP	TP + FP
	Negative	FN	TN	FN + TN
Totals		TP + FN	FP + TN	All

Example: Recall = Focus on Actual Positives

$$\text{Recall} = \frac{\textbf{TP}}{\textbf{TP + FN}} = \frac{\text{Correctly predicted positives}}{\text{All actual positives}}$$

Enhanced Confusion Matrix: Real Example

Example: Medical Diagnosis: Cancer Detection

Let's say we have 1000 patients, 100 have cancer (positive), 900 don't (negative)

Cancer Detection		Actually Has Cancer		Totals
		Yes	No	
Test Says	Positive	85	45	130
	Negative	15	855	870
Totals		100	900	1000

Enhanced Confusion Matrix: Real Example

Example: Medical Diagnosis: Cancer Detection

Let's say we have 1000 patients, 100 have cancer (positive), 900 don't (negative)

Cancer Detection		Actually Has Cancer		Totals
		Yes	No	
Test Says	Positive	85	45	130
	Negative	15	855	870
Totals		100	900	1000

Enhanced Confusion Matrix: Real Example

Example: Medical Diagnosis: Cancer Detection

Let's say we have 1000 patients, 100 have cancer (positive), 900 don't (negative)

Cancer Detection		Actually Has Cancer		Totals
		Yes	No	
Test Says	Positive	85	45	130
	Negative	15	855	870
Totals		100	900	1000

Precision vs Recall: The Trade-off

Important: Understanding the Trade-off

In most real systems, there's a trade-off between Precision and Recall!

Precision vs Recall: The Trade-off

Important: Understanding the Trade-off

In most real systems, there's a trade-off between Precision and Recall!

Definition: High Precision

Conservative model:

Precision vs Recall: The Trade-off

Important: Understanding the Trade-off

In most real systems, there's a trade-off between Precision and Recall!

Definition: High Precision

Conservative model:

- Only predicts positive when very sure

Precision vs Recall: The Trade-off

Important: Understanding the Trade-off

In most real systems, there's a trade-off between Precision and Recall!

Definition: High Precision

Conservative model:

- Only predicts positive when very sure
- Few false alarms (low FP)

Precision vs Recall: The Trade-off

Important: Understanding the Trade-off

In most real systems, there's a trade-off between Precision and Recall!

Definition: High Precision

Conservative model:

- Only predicts positive when very sure
- Few false alarms (low FP)
- But might miss

Precision vs Recall: The Trade-off

Important: Understanding the Trade-off

In most real systems, there's a trade-off between Precision and Recall!

Definition: High Precision

Conservative model:

- Only predicts positive when very sure
- Few false alarms (low FP)
- But might miss

Precision vs Recall: The Trade-off

Important: Understanding the Trade-off

In most real systems, there's a trade-off between Precision and Recall!

Definition: High Precision

Conservative model:

- Only predicts positive when very sure
- Few false alarms (low FP)
- But might miss

Definition: High Recall

Sensitive model:

Precision vs Recall: The Trade-off

Important: Understanding the Trade-off

In most real systems, there's a trade-off between Precision and Recall!

Definition: High Precision

Conservative model:

- Only predicts positive when very sure
- Few false alarms (low FP)
- But might miss

Definition: High Recall

Sensitive model:

- Tries to catch all positive cases

Precision vs Recall: The Trade-off

Important: Understanding the Trade-off

In most real systems, there's a trade-off between Precision and Recall!

Definition: High Precision

Conservative model:

- Only predicts positive when very sure
- Few false alarms (low FP)
- But might miss

Definition: High Recall

Sensitive model:

- Tries to catch all positive cases
- Few missed cases (low FN)

Precision vs Recall: The Trade-off

Important: Understanding the Trade-off

In most real systems, there's a trade-off between Precision and Recall!

Definition: High Precision

Conservative model:

- Only predicts positive when very sure
- Few false alarms (low FP)
- But might miss

Definition: High Recall

Sensitive model:

- Tries to catch all positive cases
- Few missed cases (low FN)
- But many false

Precision vs Recall: The Trade-off

Important: Understanding the Trade-off

In most real systems, there's a trade-off between Precision and Recall!

Definition: High Precision

Conservative model:

- Only predicts positive when very sure
- Few false alarms (low FP)
- But might miss

Definition: High Recall

Sensitive model:

- Tries to catch all positive cases
- Few missed cases (low FN)
- But many false

Accuracy Metrics: F-Score

		Ground Truth	
		Yes	No
Predicted	Yes	True Positive	False Positive
	No	False Negative	True Negative

$$\mathrm{F}-\text{ Score} = \frac{2\times \text{ Precision } \times \text{ Recall}}{\text{Precision } + \text{ Recall}}$$

Accuracy Metrics: Matthew's Correlation Coefficient

		Ground Truth	
		Yes	No
Predicted	Yes	True Positive	False Positive
	No	False Negative	True Negative

Matthew's correlation coefficient =

$$\frac{\mathrm{TP}\times\mathrm{TN}-\mathrm{FP}\times\mathrm{FN}}{\sqrt{(\mathrm{TP}+\mathrm{FP})(\mathrm{TP}+\mathrm{FN})(\mathrm{TN}+\mathrm{FP})(\mathrm{TN}+\mathrm{FN})}}$$

Accuracy Metrics: Example

For the data given below, calculate:

$$\begin{array}{ll} & \text{G.T. Positive} \quad \text{G.T. Negative} \\ \begin{array}{l} \text{Pred Positive} \\ \text{Pred Negative} \end{array} & \begin{pmatrix} 90 & 4 \\ 1 & 1 \end{pmatrix} \end{array}$$

Precision = ?
Recall = ?
F-Score = ?
Matthew's Coeff. = ?

Accuracy Metrics: Answer

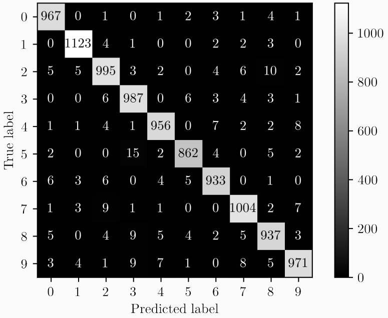
For the same data

$$\begin{array}{ccc} & \text{G.T. Positive} & \text{G.T. Negative} \\ \begin{array}{r}\text{Pred Positive} \\ \text{Pred Negative}\end{array} & \multicolumn{2}{c}{\left(\begin{array}{cc} 90 & 4 \\ 1 & 1 \end{array}\right)} \end{array}$$

Precision = $\frac{90}{94}$
Recall = $\frac{90}{91}$
F-Score = 0.9524
Matthew's Coeff. = 0.14

Confusion Matrix for multi-class classification

Notebook: confusion-mnist.html



Pop Quiz #4

Quick Quiz 4

Why might Mean Absolute Error (MAE) be preferred over Mean Squared Error (MSE)?

- a) MAE is always smaller

Pop Quiz #4

Quick Quiz 4

Why might Mean Absolute Error (MAE) be preferred over Mean Squared Error (MSE)?

- a) MAE is always smaller
- b) MAE is less sensitive to outliers

Pop Quiz #4

Quick Quiz 4

Why might Mean Absolute Error (MAE) be preferred over Mean Squared Error (MSE)?

- a) MAE is always smaller
- b) MAE is less sensitive to outliers
- c) MAE is easier to compute

Pop Quiz #4

Quick Quiz 4

Why might Mean Absolute Error (MAE) be preferred over Mean Squared Error (MSE)?

- a) MAE is always smaller
- b) MAE is less sensitive to outliers
- c) MAE is easier to compute
- d) MAE works only for classification

Pop Quiz #4 - Answer

Answer: b) MAE is less sensitive to outliers since it doesn't square the errors!

Metrics for Regression: MSE & MAE

Prediction $(\hat{y})$	Ground Truth $(\mathbf{y})$
$\begin{pmatrix} 10 \\ 20 \\ 30 \\ 40 \\ 50 \end{pmatrix}$	$\begin{pmatrix} 20 \\ 30 \\ 40 \\ 50 \\ 60 \end{pmatrix}$

$$\text{Mean Squared Error (MSE)} = \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n}$$

$$\text{Root Mean Square Error (RMSE)} = \sqrt{\text{MSE}}$$

Accuracy Metrics: MAE & ME

$$\text{Prediction } (\hat{y}) \quad \begin{pmatrix} 10 \\ 20 \\ 30 \\ 40 \\ 50 \end{pmatrix} \qquad \text{Ground Truth} \quad \begin{pmatrix} 20 \\ 30 \\ 40 \\ 50 \\ 60 \end{pmatrix}$$

$$\text{Mean Absolute Error (MAE)} = \frac{\sum_{i=1}^{n} |\hat{y}_i - y_i|}{n}$$

$$\text{Mean Error} = \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)}{n}$$

Accuracy Metrics: MAE & ME

Prediction $(\hat{y})$

$$\begin{pmatrix} 10 \\ 20 \\ 30 \\ 40 \\ 50 \end{pmatrix}$$

Ground Truth

$$\begin{pmatrix} 20 \\ 30 \\ 40 \\ 50 \\ 60 \end{pmatrix}$$

$$\text{Mean Absolute Error (MAE)} = \frac{\sum_{i=1}^{n} |\hat{y}_i - y_i|}{n}$$

$$\text{Mean Error} = \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)}{n}$$

Is there any downside with using mean error?

Accuracy Metrics: MAE & ME

$$\text{Prediction } (\hat{y}) \quad \begin{pmatrix} 10 \\ 20 \\ 30 \\ 40 \\ 50 \end{pmatrix} \qquad \text{Ground Truth} \quad \begin{pmatrix} 20 \\ 30 \\ 40 \\ 50 \\ 60 \end{pmatrix}$$

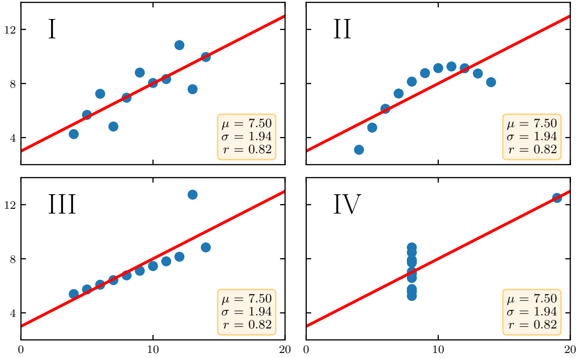
$$\text{Mean Absolute Error (MAE)} = \frac{\sum_{i=1}^{n} |\hat{y}_i - y_i|}{n}$$

$$\text{Mean Error} = \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)}{n}$$

Is there any downside with using mean error?
Errors can get cancelled out

The Importance of Plotting

Notebook: anscombe.html



Anscombe's Quartet

Dummy Baselines

Notebook: dummy-baselines.html

The Importance of Plotting

Property	Value	Across datasets
mean(X)	9	exact
mean(Y)	7.5	up to 3 decimal pla
Linear regression line	y = 3.00 + 0.500x	up to 2 decimal pla

Pop Quiz #5

Quick Quiz 5

For imbalanced datasets, which metrics should you prioritize over accuracy?

- a) Only precision

Pop Quiz #5

Quick Quiz 5

For imbalanced datasets, which metrics should you prioritize over accuracy?

- a) Only precision
- b) Only recall

Pop Quiz #5

Quick Quiz 5

For imbalanced datasets, which metrics should you prioritize over accuracy?

- a) Only precision
- b) Only recall
- c) Precision, recall, and F1-score

Pop Quiz #5

Quick Quiz 5

For imbalanced datasets, which metrics should you prioritize over accuracy?

- a) Only precision
- b) Only recall
- c) Precision, recall, and F1-score
- d) Only confusion matrix

Pop Quiz #5 - Answer

Answer: c) Precision, recall, and F1-score give a more complete picture!

Key Takeaways

Key Points

- **ML vs Traditional Programming:** ML learns rules from data, traditional programming uses predefined rules

Key Takeaways

Key Points

- **ML vs Traditional Programming:** ML learns rules from data, traditional programming uses predefined rules
- **Features matter:** Choose meaningful features, avoid arbitrary identifiers

Key Takeaways

Key Points

- **ML vs Traditional Programming:** ML learns rules from data, traditional programming uses predefined rules
- **Features matter:** Choose meaningful features, avoid arbitrary identifiers
- **Classification vs Regression:** Discrete outputs vs continuous outputs

Key Takeaways

Key Points

- **ML vs Traditional Programming:** ML learns rules from data, traditional programming uses predefined rules
- **Features matter:** Choose meaningful features, avoid arbitrary identifiers
- **Classification vs Regression:** Discrete outputs vs continuous outputs
- **Accuracy isn't everything:** For imbalanced data, use precision, recall, F1-score

Key Takeaways

Key Points

- **ML vs Traditional Programming:** ML learns rules from data, traditional programming uses predefined rules
- **Features matter:** Choose meaningful features, avoid arbitrary identifiers
- **Classification vs Regression:** Discrete outputs vs continuous outputs
- **Accuracy isn't everything:** For imbalanced data, use precision, recall, F1-score
- **Visualization is crucial:** Always plot your data (Anscombe's Quartet lesson)

Key Takeaways

Key Points

- **ML vs Traditional Programming:** ML learns rules from data, traditional programming uses predefined rules
- **Features matter:** Choose meaningful features, avoid arbitrary identifiers
- **Classification vs Regression:** Discrete outputs vs continuous outputs
- **Accuracy isn't everything:** For imbalanced data, use precision, recall, F1-score
- **Visualization is crucial:** Always plot your data (Anscombe's Quartet lesson)
- **Use baselines:** Simple baseline models help validate your approach

Summary: Evaluation Metrics

Task	Common Metrics	When to Use
Classification	Accuracy, Precision, Recall, F1 Confusion Matrix	Balanced/Imb Multi-class pr
Regression	MSE, RMSE, MAE Mean Error	Continuous p Check for bia

Remember: Choose metrics based on your problem's characteristics and business requirements!

Colorbox Examples - Definition & Example

Definition: Machine Learning

A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E.

Example: Classification Task

Predicting whether an email is spam or not spam based on features like sender, subject line, and content.

Colorbox Examples - Alert & Theorem

Important: Common Mistake

Never use accuracy alone for imbalanced datasets! A classifier that always predicts the majority class can have high accuracy but be completely useless.

Theorem: Bayes' Theorem

For events A and B: $P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$